CLONEID: A Framework for Monitoring and Steering Subclonal Dynamics

Thomas Veith[+2], Richard Beck[+2], Vural Tagal[+2], Tao Li[+2], Saeed Alahmari[1], Jackson Cole[2], Daniel Hannaby , Xiaoqing Yu[3], Konstantin Maksin[8], Andrew Schultz[2], HoJoon Lee[5], Issam ElNaqa[4], Steven Eschrich[3], Janine Lupo[7], Hanlee P. Ji[5], Aaron Diaz[6], and Noemi Andor[2]

[1]Department of Computer Science, Najran University, Najran 66462, Saudi Arabia
[2]Department of Integrated Mathematical Oncology, H Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA
[3]Department of Biostatistics and Bioinformatics, H Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA
[4]Department of Machine Learning, H Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA
[5]Department of Medicine, Stanford University, CA, USA
[6]Department of Epidemiology & Biostatistics, University of California San Francisco, CA, USA
[7]Department of Radiology & Biomedical Imaging, University of California San Francisco, CA, USA
[8]The University of Commerce and Services, Medical Department (WSHIU), Poznan, Poland

[+]Authors contributed equally to this work [*]Corresponding author: noemi.andor@gmail.com

Abstract

Understanding how genetic and phenotypic diversity emerges and evolves within cancer cell populations is a fundamental challenge in cancer biology. CLONEID is a novel framework designed to organize and analyze clone-specific measures as structured data. By integrating genotypic and phenotypic experimental data, CLONEID facilitates hypothesis-driven and hypothesis-generating research in cancer biology. This article outlines the development, utility, and applications of CLONEID, emphasizing its role in overcoming challenges in data reproducibility, mathematical modeling, and experimental design. A webportal to the CLONEID database is available at https://dev.cloneid.org.

1 Introduction

Cancer is a dynamic, phenotypically diverse disease driven by underlying genetic, epigenetic, and transcriptional variation. Within a single tumor, distinct subpopulations of cells emerge with different genotypes and gene expression programs [1]. This *intra-tumor heterogeneity* (ITH) is evident in live-cell and clinical

imaging studies: intravital microscopy and functional medical imaging have directly visualized coexisting phenotypic differences, such as variations in proliferation rates, invasive behavior, metabolic activity, and drug responsiveness among cells in the same tumor[2, 3]. These observations highlight remarkable plasticity and adaptability in cancer, as genetically similar cells can adopt divergent states in response to microenvironmental cues or therapy pressure.

Concurrently, single-cell sequencing technologies have mapped the genetic and transcriptional landscape of ITH at high resolution. Single-cell DNA sequencing reveals that most tumors comprise multiple clonal lineages harboring distinct mutation and copy number aberration profiles, reflecting branched evolutionary histories within the cancer[4]. Likewise, single-cell RNA sequencing has uncovered diverse transcriptional cell states within individual tumors – for example, coexisting stem-like, differentiated, stress-resistant, and proliferative subpopulations – even when cells share the same underlying mutations[5].

Numerous computational tools have been developed to quantify ITH using data from bulk-, single-cell and spatial omics technologies [6, 7, 8, 5, 4]. These tools facilitate the analysis of genetic, epigenetic, and transcriptional diversity within tumors, enhancing our understanding of cancer complexity. Despite these advances, there is currently no unified framework to compare and track the outputs of different ITH quantification methods across time or across studies. Moreover, a significant gap remains between imaging and sequencing modalities: genomic sequencing datasets often lack crucial metadata such as lineage history, sampling time, or growth conditions, while imaging experiments rarely include direct genetic or transcriptomic readouts. This disconnect hampers efforts to connect phenotypic traits observed in vivo with the underlying molecular clones characterized by sequencing and other high-throughput assays.

We introduce CLONEID to help bridge these gaps. We begin by outlining how CLONEID's infrastructure facilitates collection, connection, and analysis of genotypic and phenotypic data from in-vitro experiments, in-vivo and clinical studies, across laboratories and timeframes. We also introduce the first database compiled using CLONEID, detailing its structure and contents. Next, we demonstrate CLONEID's utility through three distinct use cases, categorized into two broad applications: hypothesis-driven mathematical modeling and hypothesis-generating association studies or data mining. Each use case is supported by a multi-modal dataset integrated into the CLONEID database, comprising both genotypic and phenotypic data. This framework aims to support the development of mathematical models and foster reproducibility in biological research.

2 Methods and Framework

CLONEID is implemented as a three-tier framework consisting of: (i) a user interface for data visualization and analysis, (ii) a core functionality layer implemented in Java, and (iii) an SQL database for data storage and organization. The CLONEID database provides a centralized, structured resource that can store clone-specific data, including size, expression, single-cell DNA/RNA sequencing profiles, and lineage information.

These components are structured into two modules: the phenotype module, which handles image analysis and integration of metainformation, and the multi-omics module, which manages data on the clonal composition of a sample or biopsy as derived from high-throughput assays.

2.1 Phenotype module monitors growth dynamics from routine image analysis

Mounting evidence shows that every cell division can yield mutations that help cells adapt to their environment [9, 10, 11, 12, 13, 14], underscoring the need to monitor real-time changes. Traditional software for cell culture management (e.g., CultureTrax, Benchling) offers only isolated snapshots [15, 16], limiting insights into how phenotypes emerge, persist, or transition. Imaging data collected in various settings—including clinical trials, in-vivo experiments, and in-vitro studies—typically remain disconnected from biopsy-derived genomic and high-throughput omics analyses, particularly with respect to clonality.

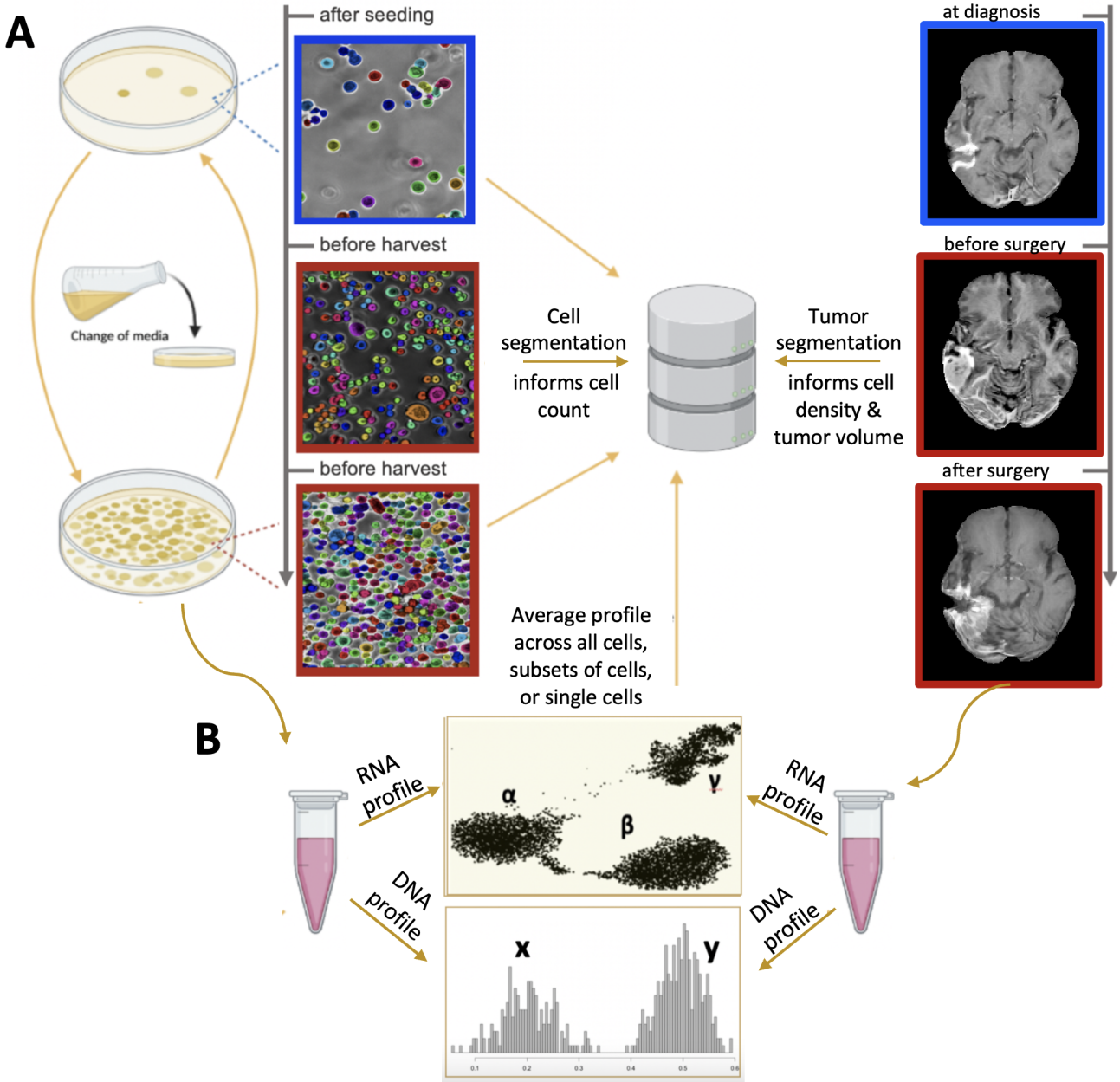

Figure 1: CLONEID workflow and central components. (**A-B**) CLONEID's database is designed to monitor phenotypic (A) and genotypic (B) information. (**A**) *Seedings*, *harvests*, *injections*, *resections*, and other sample-related events are recorded in real-time. Each is linked to time-stamped phenotypic data such as cell counts, growth conditions, or tumor volume. Shown are brightfield images from an in-vitro gastric cancer experiment (NUGC-4, segmented with Cellpose) and clinical radiology data integrated through the same interface. (**B**) scRNA/DNA-seq data identifies clones by copy number state and quantifies gene expression. Recursive clonal membership structure allows for analysis of all cells, subsets, or single cells. Panels A and B created with BioRender.com.

CLONEID's phenotype module unifies these data through an event-based capture system, recording in-vitro, in-vivo, or clinical events in real time (Figure 1A). For in-vitro studies, events such as cell *seeding* and *harvesting* are logged systematically. In-vivo procedures like cell injections (*inject*) are documented along with mouse-specific metadata. Clinical events, including tissue resection (*resect*), diagnosis (*diagnose*), and follow-up visits (*follow_up*), are also captured. Critically, CLONEID links these temporally organized events to downstream analytic modules, which automatically extract and store image-derived sample characteristics such as cell segmentation, media conditions, growth rates, and cell counts. These

interfaces ensure that time-sensitive information is accurately stored and linked to relevant experimental events. At its core is a *Passaging* table that tracks every propagation or sampling event—whether an in-vitro passage or a clinical biopsy—capturing sample lineage, event type, date, and other relevant metrics. By linking each time-resolved record to its originating sample or lineage, this table enables users to reconstruct detailed cell ancestries, track how long cells have remained in specific environments, and systematically connect phenotypic observations to genotypic profiles across different stages of evolution. Unlike conventional systems that treat samples as static identifiers, CLONEID captures the full temporal context of each biological entity.

2.2 Multi-omics module monitors changes in clonal composition

While several algorithms have been developed to infer clonal composition, no framework exists that systematically stores clone-specific measures alongside imaging and phenotypic metadata. CLONEID's multi-omics module provides a structured and scalable database for storing and analyzing clone-specific information, including clone size, gene expression, and copy number profiles across multiple assays and resolutions (Figure 1B).

While CLONEID currently offers built-in support for preexisting segmentation tools (e.g., Cellpose) for deriving morphological features from images, it does not derive omics-based features (e.g., copy number or gene expression values) due to the diversity of technologies and tools involved. Instead, it is designed to ingest outputs from widely-used tools and integrate them with rich phenotypic and temporal metadata. By contextualizing user-supplied omics data within a standardized schema, CLONEID enables downstream reasoning, quality control, and the design of evolutionary or statistical models. This decoupling of data derivation and integration allows users to flexibly choose their preferred analysis pipelines while still benefiting from a centralized framework for interpretation and hypothesis generation.

The core of the multi-omics module is the `Perspective` table, which holds genomic information at different levels of cellular resolution: (i) averaged over all profiled cells, (ii) at the level of defined subpopulations (e.g., clones), or (iii) for individual cells (e.g., from scRNA-seq or scDNA-seq). This layered design enables both computational and mathematical biologists to trace not only the phenotype of profiled cells, but also that of their ancestors and the duration each has spent in a particular environment. The `Perspective` table is linked to CLONEID's phenotype module, allowing users to view and analyze genotype-phenotype relationships across time and conditions.

To support clonal tracking, CLONEID uses a second table, `Identity`, which synthesizes information across perspectives to approximate each clone's genomic identity. This schema allows users to reason about clonal structure across samples, identify shifts in clonal composition, and evaluate the consistency and quality of clone detection. Crucially, it supports evolutionary modeling by enabling users to trace clonal histories, relate changes in environment to selection, and visualize how clonal architecture unfolds over time.

To capture the relationship between cell-level and clone-level data, CLONEID expects inputs at up to two levels of resolution: first, the properties of each clone (e.g., mean expression or copy number profile), and second, the member cells associated with that clone. This design ensures accurate tracking of clonal dynamics over time, while offering a standardized foundation for downstream modeling.

2.3 Data Ingestion and Annotation Workflow

The CLONEID system provides a structured workflow for data ingestion, ensuring comprehensive capture and annotation of experimental information at multiple biological scales. The ingestion process begins by registering each new specimen or cell lineage into the database, establishing a unique and persistent identity. Subsequently, the system tracks relevant experimental events, such as passaging, media changes, or treatments, along with their precise timings.

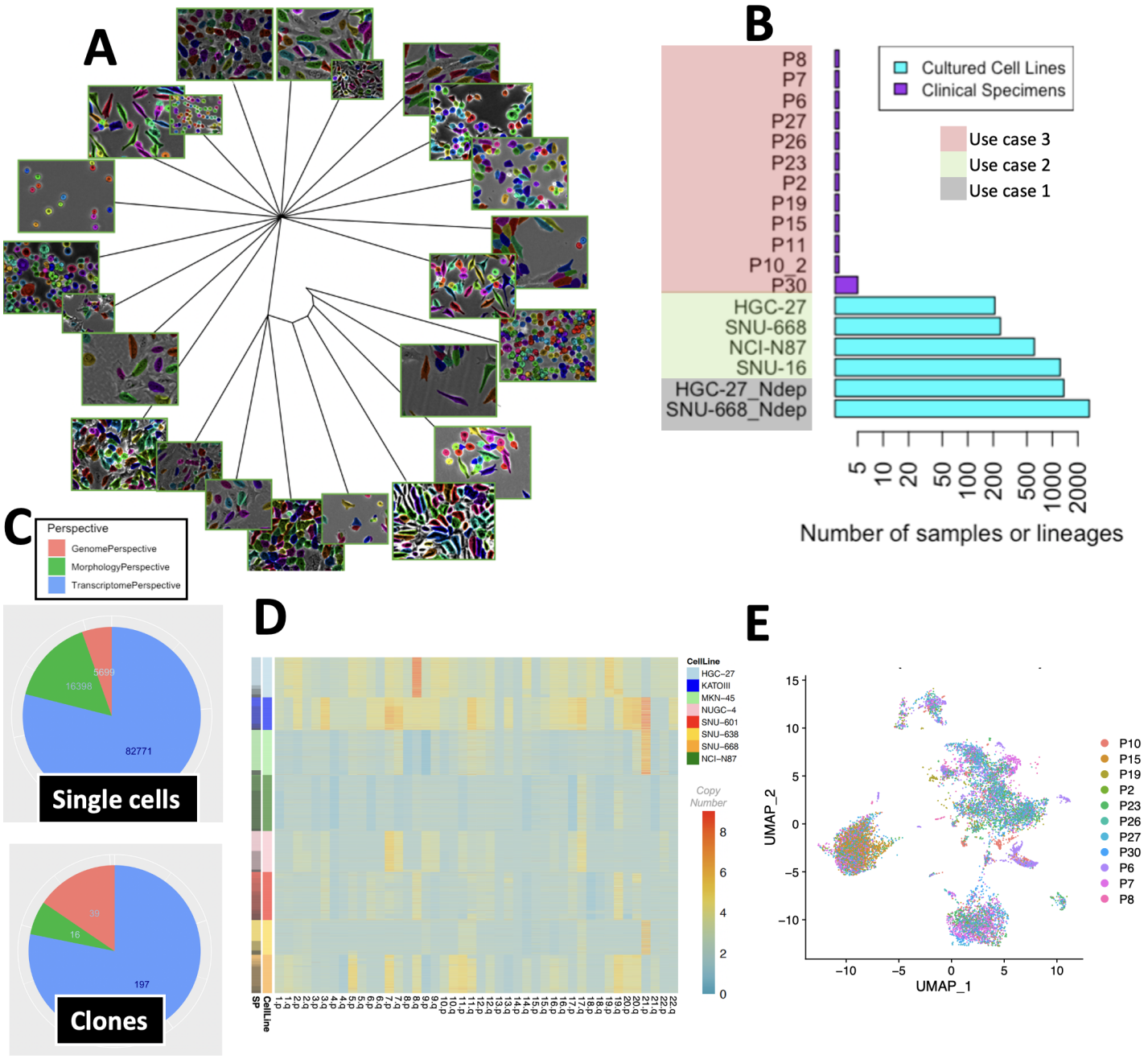

Figure 2: Database content. (A) Pedigree of the gastric cancer cell line HGC-27, with 23 lineages. Representative cells for each lineage are shown along with their segmentation masks. Pedigree trees in our CLONEID database each have between 2 - 2,726 lineages (mean=165; median=9), all accompanied by high-quality images and segentation masks. (B) Number of lineages per case represented in the Passaging table, color coded by use case and origin. (C) Number of clones (bottom) and single cells (top) represented in the Perspective table. (D) Chromosome-arm level copy number profiles across 5,779 cells from eight cell lines included in the database displayed as heatmap. (E) Gene expression profiles derived from 59,251 scRNA-sequenced glioma cells across eleven GBM patients, displayed as UMAP.

When data acquisition events occur—such as microscopy imaging, live-cell tracking, digital pathology, or multi-omics profiling—they are registered in the system and linked directly to the specimen's position in its lineage tree. This ensures that data points are contextualized within the clone's evolutionary and experimental history.

For single-cell and spatial omics data, CLONEID supports hierarchical annotation at multiple levels: (i) Subpopulation level, reflecting shared clonal identity or state-specific expression profiles and (ii) Cellular level, allowing the resolution of heterogeneity within clones or tissue regions. These annotations are stored in modular data layers (e.g., GenomePerspective, TranscriptomePerspective) and linked via unique specimen and cell identifiers to ensure integrative analysis across data types and time points.

2.4 Overview of Database Content

As of this publication, the CLONEID database contains over 5,000 lineages/specimens across 18 clinical and in-vitro cases, with data types ranging from single-cell DNA and RNA sequencing to morphological features. CLONEID has recorded over 30,000 microscopy, digital pathology, and medical images, all embedded in the context of large, branching pedigree trees (Figure 2A).

In addition to image-based and phenotypic data, the multi-omics module of CLONEID enables integrated analysis of genomic and transcriptomic profiles. A key feature of this module is that each data layer accommodates varying resolutions—from bulk averages to single-cell measurements—already described in a previous section. Building on this capability, CLONEID currently includes omics data from 104,896 individual cells and 252 distinct clones, organized into modular layers such as *GenomePerspective* and *TranscriptomePerspective* (Figure 2B). These layers are further linked to clonal identities via the *Identity* and *Perspective* tables, enabling downstream analyses of clonal selection, tumor evolution, and genotype-phenotype relationships within the full temporal and lineage context of each experiment.

3 Use Cases

We present three use cases demonstrating the utility of CLONEID. Two leverage CLONEID data to test hypotheses in vitro or in clinical settings, while the other one is a hypothesis-generating association study using in vitro data. All use cases integrate both genotypic and phenotypic data stored in the CLONEID database.

Use Case 1: Tracking Clonal Adaptation Under r/K-Selection in Gastric Cancer Cell Lines

Type of Use Case: Mechanistic, In-Vitro

We tested the impact of ecological selective pressures on a gastric cancer cell line (SNU-668). Evolution experiments were conducted by repeatedly passaging cells under low or high confluence conditions for over six months. Under r-selection, cells were seeded at low density (Fig. 3B) and allowed to grow exponentially before being re-plated, while K-selection involved high-density seeding with transfer after growth plateaued. Each lineage evolved for more than 40 passages under its respective regime. To assess phenotypic consequences of this long-term selection, we measured growth rates across density conditions (Fig. 3C). As expected, lineages evolving under r-selection exhibited significantly higher growth rates than those under K-selection, as they had more space to grow.

After being cultured under these conditions for 27 and 55 passages, for K-selection and r-selection respectively, samples from each of the six test condition replicates were sent for karyotyping along with two samples from the parental lineage as controls (28 K-selected cells, 61 r-selected, and 32 control). Clustering these 121 cells by karyotype we see clusters forming by condition: most cells from the test conditions have regained a copy of chromosome 21, relative to the control lineages (Fig. 3E). In addition

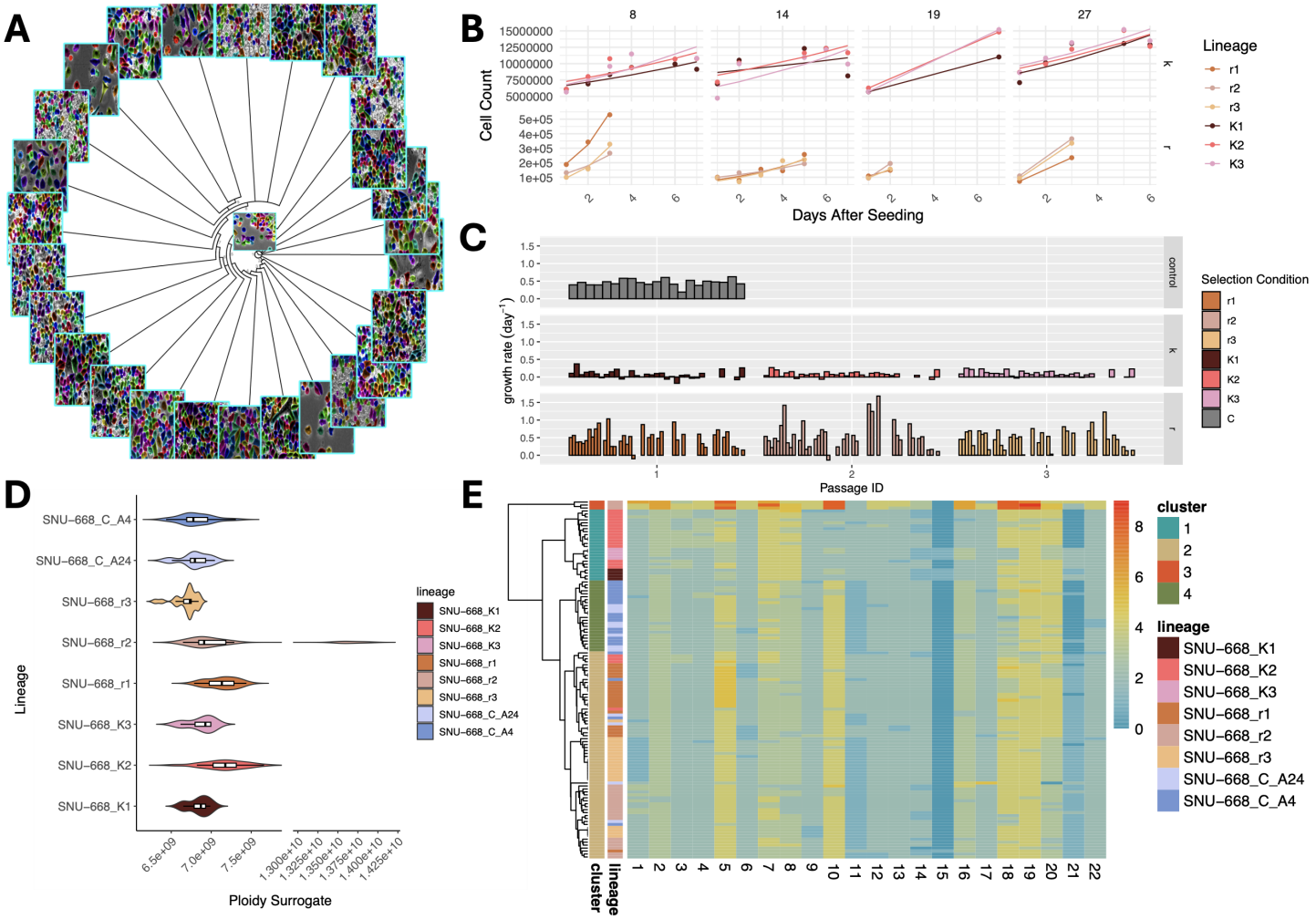

Figure 3: CLONEID reveals the impact of density-dependent selection on cancer evolution. (A) Pedigree of a subset of the K-selected SNU-668 cells, showing 9 lineages. Representative cells for each lineage are shown along with their segmentation masks. (B) Daily cell counts (datapoints) over time and fitted growth curves are shown over multiple passages (columns) for six replicates of adaptation to sparse (r-selection) or dense (K-selection) population conditions (rows and color legend). (C) Changes in growth rate (y-axis) over time (x-axis) for all replicates and passages are shown. All information displayed in (B,C) was retrieved directly by mining the CLONEID database. (D-E) Karyotype evolution can facilitate adaptation to density conditions in SNU-668. (D) Two control (blue) and six test samples from the SNU-668 cell line were karyotyped at late passages of their adaptation to their specific density conditions. Cumulative area of all segmented chromosomes was used as surrogate of ploidy (x-axis). Whole genome doubling observed in the r2 lineage. (E) Clustering cells (y-axis) according to their chromosome level copy number (x-axis) shows that cells cluster according to their density conditions (color bar).

we note several differences between r-selected and K-selected cells, including loss of chromosomes 10 and 14 in K cells (Fig. 3E). Karyotyping revealed that r-selection also enriched for tetraploidized cells (Fig. 3D-E), indicating that whole genome doubling may be favored in early, low-competition environments. These data suggest that population density and resource availability are key features of early tumorigenesis that can shape the evolutionary trajectory of cancer cells.

Experiments with gastric cancer cell lines under varying population densities demonstrated the utility of CLONEID in monitoring adaptation. Data collected informed decisions about further profiling, such as karyotyping, and enabled the fitting of population growth models to observe phenotypic changes.

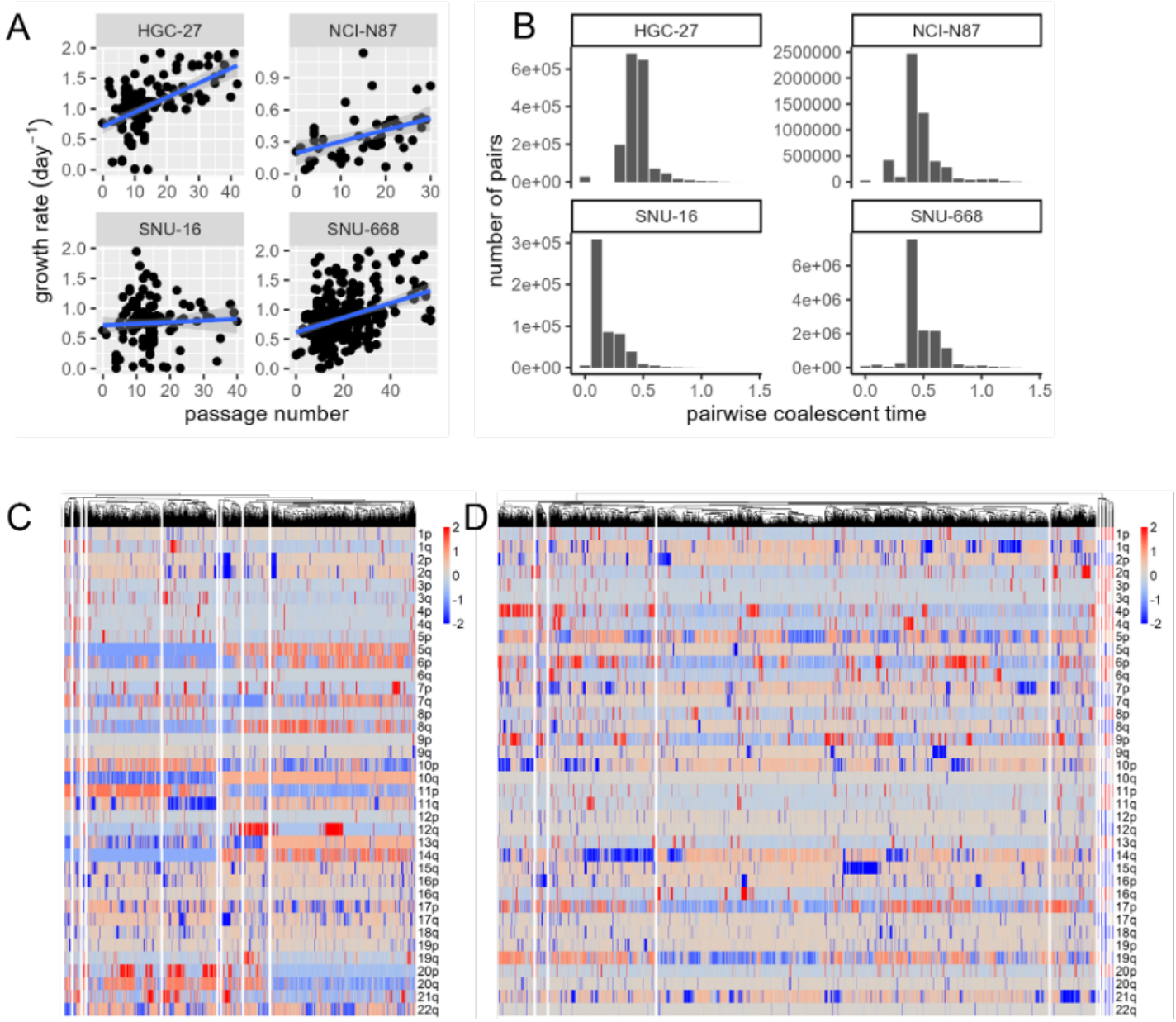

Figure 4: Predicting cellular plasticity from a snapshot of the transcriptome. (A) Passaging data were used to reconstruct lineage networks, and logistic growth models were fit to cell count trajectories from passages with at least three time points to estimate intrinsic growth rates. These growth rates were plotted against passage number to reveal temporal changes in proliferative dynamics. (C,D) CONICSmat was applied to scRNA-seq data to infer large-scale copy-number variations from normalized expression matrices. After scaling and thresholding the data, hierarchical clustering produced heatmaps with annotated cluster IDs. (B) Subsequent dendrogram analysis enabled computation of coalescent times to summarize clonal evolution. Lineages with growth rates that increase across successive passages tend to exhibit higher coalescent times in the CNV-derived dendrograms, indicating higher clonal heterogeneity may permit adaptation.

Use Case 2: Predicting Cellular Plasticity from a Snapshot of the Transcriptome

Type of Use Case: Data Mining/ML, In-Vitro

Four gastric cell lines in the CLONEID database (HGC-27, NCI-N87, SNU-16, and SNU-668) have evolved over 40 or more passages, representing months of continuous culture. Near-daily images of the growing cells were taken and analyzed to quantify cell numbers. A logistic growth model was fit to these cell count data to investigate changes in growth rate over time. In three of the four cell lines, significant increases in growth rate were observed, indicating adaptation to culture conditions (Figure 4).

To predict adaptation from transcriptomic data, copy-number-based phylogenetic trees were constructed using CONICs [17]. For each tree, pairwise coalescence times were computed for sequenced cells. For SNU-16, which did not display strong growth rate changes, coalescence times aligned closely with neutral evolution predictions. Its phylogenetic tree was more balanced compared to the adapting cell lines, as shown in Figure 4.

These results demonstrate that phylogenetic trees constructed from single snapshots of transcriptomic data stored with CLONEID, can predict selection and growth rate changes in evolving cancer cell lines. These analyses highlight the potential of CLONEID to drive hypothesis-generating research.

Use Case 3: Inferring karyotype fitness landscapes from patient data

Type of Use Case: Mechanistic, Retrospective

Accumulating evidence highlights that the phenotypic impact of copy number alterations (CNAs) is highly context-dependent, with their effects influenced significantly by other concurrent genomic events. For instance, an analysis of chromosome-arm aneuploidies across 6,977 tumors identified numerous pairs of synergistically co-occurring aneuploidies that served as prognostic biomarkers, even though individual alterations lacked predictive power when considered independently of their broader karyotypic context [18]. We used CLONEID to record MRI imaging (Fig. 5A) and karyotype data (Fig. 5B) from a cohort of 11 glioblastoma (GBM) patients with matched primary and recurrent tumor specimens [19, 20]. We then applied ALFA-K to model the karyotype fitness landscape for each patient [21]. ALFA-K (Adaptive Local Fitness Landscapes for Aneuploid Karyotypes) is a computational framework designed to predict the fitness consequences of copy number alterations (CNAs) in a karyotype context-dependent manner, enabling prediction of tumor evolution from single-cell karyotype data. For instance, in patient P30, ALFA-K inferred the fitness of 28,405 unique karyotypes (Fig. 5C) based on the dynamics of 36 karyotypes observed in the input data (Fig. 5B). Cross-validation of the inferred fitness landscapes demonstrated strong concordance for 9 out of 11 cases, which were subsequently included in downstream analyses.

Across all nine landscapes, we identified 203,064 unique karyotypes (Fig. 5E), of which 9.47% (19,230 karyotypes) occurred in more than one landscape (hereafter referred to as shared karyotypes) and 0.02% (45 karyotypes) were present in all landscapes (ubiquitous karyotypes). We calculated the Pearson correlation coefficient between the fitness values of shared karyotypes across patient pairs, identifying significant positive and negative correlations in 13 and 20 out of 36 pairs, respectively (Fig. 5G; bubble chart above diagonal). Pearson correlation coefficient between the fitness values of ubiquitous karyotypes were higher (Fig. 5F), with 6 positive and 3 negative correlations (Fig. 5G; bubble chart below diagonal). This context-dependent fitness highlights the importance of further investigation into personalized factors driving variability in selection pressures among patients.

To further characterize karyotype selection in this GBM cohort, we computed the proportion of karyotypes in which a given copy number alteration (CNA) conferred a net fitness advantage (Fig. 5D). This analysis revealed that most CNAs are deleterious, while beneficial karyotypes are largely context-dependent. Overall, these findings demonstrate that karyotype fitness landscapes exhibit both patient-specific and shared selection pressures, underscoring the heterogeneous yet partially predictable nature of karyotype evolution in GBM.

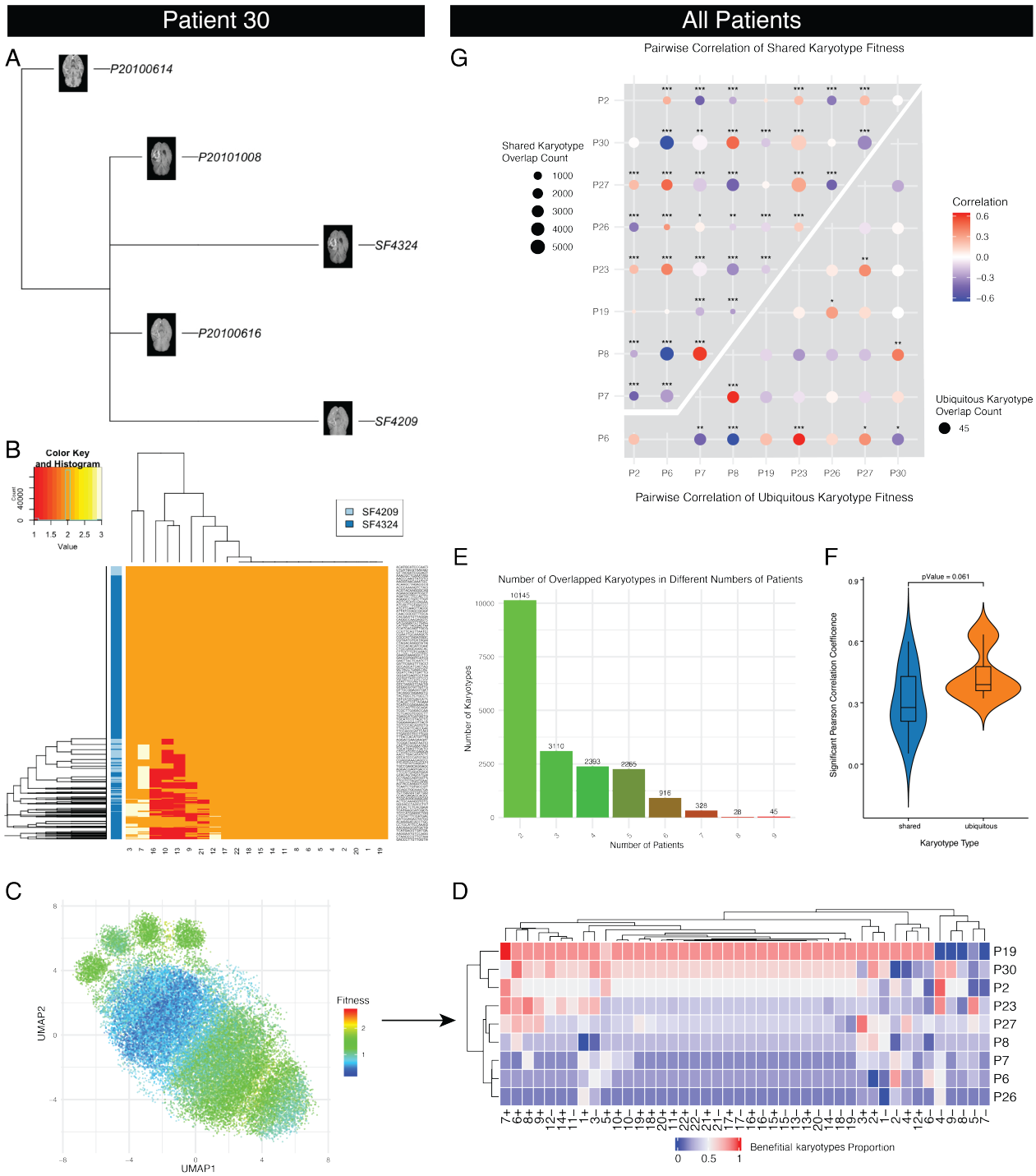

Figure 5: Multimodal analysis of a Glioblastoma patient through CLONEID informs a mathematical model of karyotype evolution. (A) MRI data from three timepoints were viewed through CLONEID's phenotype module; biopsies were taken during surgery at two timpoints (prefix SF*). (B) Karyotype composition derived from scRNA-seq of the two biopsies. 4,471 cells (rows) were sequenced in total and were used as input to a mathematical model for inferring karyotype fitness landscapes. (C) UMAP representation of 28,405 unique karyotypes that make up the charted fitness landscape for P30. Each karyotype is color-coded by its inferred fitness. (E) The histogram shows the karyotype count (Y-axis) shared across patient samples (X-axis). (F)Violin and box plots display the distribution of Pearson correlation coefficients between fitness and shared or ubiquitous karyotype groups (t-test). (G) Pairwise correlation coefficients between karyotypes shared by the given pair of patients (above diagonale) and between karyotypes shared by all patients (below diagonale). (D) Fitness consequence of copy number alterations (CNAs; x-axis) is shown for P30 along with all analyzed patients (y-axis). Entries calculated as proportion of karyotypes in C), for which a given CNA increases fitness.

4 Discussion

We present a framework for routine and systematic tracking of the pedigree of evolving cell populations over decades along with potentially changing environments (e.g. cell culture habits, therapy). Such efforts could help reveal long-term trends in the evolution of cell lines and tumors, that remain elusive at shorter time-scales. A system that tracks clones and how they change over time widens the path for genotype-phenotype mapping and discovery at large scale: e.g. in what direction do changes in culture conditions, or administration of specific therapeutics steer a population's subclone composition?

CLONEID is designed as a collaborative platform that bridges the gap between cancer biology and computational modeling, enabling seamless interaction between experimentalists and computational or mathematical biologists. Experiments to inform mathematical models are carefully planned to ensure they will provide the necessary information for modeling, a process that requires coordination and discussions between experimentalists and computational/mathematical biologists and often takes weeks per experiment. CLONEID provides a workflow to streamline this process. It ensures that the information needed for modeling is permanently stored in a centralized location with minimal extra burden on the experimentalist. For example, in addition to a cell-by-feature matrix, the lineage tracing module records time at which the underlying image was taken – a critical record for modeling that is often missing from routine experiments.

Consistency in the usage of CLONEID is key and crystallizes into several use cases: 1) Longitudinal information recorded by CLONEID informs parameters of mathematical models. This can be done with the lineage tracing module, with the multi-omics module or with both in combination depending on the hypothesis being tested with the mathematical model. 2) Genetic and morphologic features collected by the multi-omics and lineage tracing module respectively can be used for association studies investigating how genes affect cell morphology. 3) Long-term cell culture experiments where cells are challenged to adapt to a new environment (e.g. limiting Glucose availability). For such experiments, it is difficult to predict how long they will last (often several months or even years) and at what interval to sample cells for sequencing. Monitoring subtle morphological changes, such as changes in cell size, can inform experimentalists when to sample cells for sequencing. CLONEID facilitates this by automatically storing morphological features in the context of other information such as cell density, time, pedigree. 4) Routine recording of cell images over years opens the stage for deep learning applications, which require data at a scale beyond the scope of most experimental designs. Since its inception, CLONEID has stored over 30,000 microscopy images, each paired with rich metadata and embedded in the context of large pedigree trees (Figure 2A). Notably, this remarkable dataset has been generated by a single lab in just a few years. Scaling this platform to support research across entire laboratories and careers worldwide could unlock unparalleled insights and generate invaluable data for the broader scientific community.

While the primary goal of CLONEID is to provide infrastructure for efficient collaborations between experimentalists, computational biologists, and mathematicians, its design also offers a powerful means to quantify biological variability, which is a critical step toward improving reproducibility in cancer biology. More than 70% of researchers have tried and failed to reproduce another scientist's experiments, and more than half have failed to reproduce their own [22]. In oncology, 97% of clinical trials fail despite promising preclinical results, with poor authentication of biological reagents being the leading cause of irreproducibility in in-vitro and in-vivo studies. While the precise contribution of cell line heterogeneity to this problem is unknown, there is substantial evidence that it plays a significant role. For example, work by Uri Ben-David et al. (2018) demonstrated that even slight differences in media composition can drive the selection of genetically distinct subpopulations, which in turn dramatically affect drug response. CLONEID addresses this challenge not just by controlling for variability, but by making it measurable and transparent. It systematically tracks (i) the pedigree of cell lineages, (ii) the exact media composition, and (iii) the number of cell divisions—an important proxy for the evolutionary time each clone experiences. By capturing this level of granularity, CLONEID enables researchers to interpret findings in the context of known variability.

We expect CLONEID to streamline acquisition of phenotypic measurements at high temporal resolution, as a byproduct of any *in-vitro/in-vivo* experiment, without complicating experimental design. Accumulating evidence that cancer cell lines are genetically heterogeneous and dynamic [13, 9, 10, 23], is primarily seen as a tremendous inconvenience. But this paradigm shift places these model systems closer to the clinical reality of primary tumors, which also adapt and change in response to their environment. Success in the clinical setting requires an experimental setting in which our therapeutic solutions are challenged with the adaptive nature of cancer. We seek to contribute to a paradigm shift in the way research laboratories define success of pre-clinical studies, not as the maximum reduction in cancer cell population growth, but as the ability to control a population's clonal dynamics over several generations of environmental adaptation.

Acknowledgments

We thank Sue Grimes for her early contributions to the database design, which helped ensure its long-term scalability.

Funding

This work was supported by the NCI grants 1R03CA259873-01A1, 1R37CA266727-01A1 and 1R21CA269415-01A1 awarded to N. Andor. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflict of Interest

The authors declare no conflicts of interest.

References

[1] M. Greaves and C. C. Maley. Clonal evolution in cancer. Nature, 481(7381):306–313, January 2012.

[2] J. P. O'Connor, C. J. Rose, J. C. Waterton, R. A. Carano, G. J. Parker, and A. Jackson. Imaging Intratumor Heterogeneity: Role in Therapy Response, Resistance, and Clinical Outcome. Clinical cancer research : an official journal of the American Association for Cancer Research, 21(2):249–257, January 2015. PMID: 25421725.

[3] A. Kreso, C. A. O'Brien, P. van Galen, O. I. Gan, F. Notta, A. M. K. Brown, K. Ng, J. Ma, E. Wienholds, C. Dunant, A. Pollett, S. Gallinger, J. McPherson, C. G. Mullighan, D. Shibata, and J. E. Dick. Variable clonal repopulation dynamics influence chemotherapy response in colorectal cancer. Science (New York, N.Y.), 339(6119):543–548, February 2013. PMID: 23239622.

[4] N. Navin, J. Kendall, J. Troge, P. Andrews, L. Rodgers, J. McIndoo, K. Cook, A. Stepansky, D. Levy, D. Esposito, L. Muthuswamy, A. Krasnitz, W. R. McCombie, J. Hicks, and M. Wigler. Tumour evolution inferred by single-cell sequencing. Nature, 472(7341):90–94, April 2011.

[5] A. P. Patel, I. Tirosh, J. J. Trombetta, A. K. Shalek, S. M. Gillespie, H. Wakimoto, D. P. Cahill, B. V. Nahed, W. T. Curry, R. L. Martuza, D. N. Louis, O. Rozenblatt-Rosen, M. L. Suvà, A. Regev, and B. E. Bernstein. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science (New York, N.Y.), 344(6190):1396–1401, June 2014. PMID: 24925914.

[6] C. Ma, M. Balaban, J. Liu, S. Chen, M. J. Wilson, C. H. Sun, L. Ding, and B. J. Raphael. Inferring allele-specific copy number aberrations and tumor phylogeography from spatially resolved transcriptomics. Nature Methods, 21(12):2239–2247, December 2024.

[7] S. Zaccaria and B. J. Raphael. Accurate quantification of copy-number aberrations and whole-genome duplications in multi-sample tumor sequencing data. Nature Communications, 11(1):4301, September 2020.

[8] A. Roth, J. Khattra, D. Yap, A. Wan, E. Laks, J. Biele, G. Ha, S. Aparicio, A. Bouchard-Côté, and S. P. Shah. PyClone: statistical inference of clonal population structure in cancer. Nature Methods, 11(4):396–398, April 2014.

[9] U. Ben-David, B. Siranosian, G. Ha, H. Tang, Y. Oren, K. Hinohara, C. A. Strathdee, J. Dempster, N. J. Lyons, R. Burns, A. Nag, G. Kugener, B. Cimini, P. Tsvetkov, Y. E. Maruvka, R. O'Rourke, A. Garrity, A. A. Tubelli, P. Bandopadhayay, A. Tsherniak, F. Vazquez, B. Wong, C. Birger, M. Ghandi, A. R. Thorner, J. A. Bittker, M. Meyerson, G. Getz, R. Beroukhim, and T. R. Golub. Genetic and transcriptional evolution alters cancer cell line drug response. Nature, 560(7718):325–330, August 2018. PMID: 30089904.

[10] U. Ben-David, G. Ha, Y.-Y. Tseng, N. F. Greenwald, C. Oh, J. Shih, J. M. McFarland, B. Wong, J. S. Boehm, R. Beroukhim, and T. R. Golub. Patient-derived xenografts undergo mouse-specific tumor evolution. Nature Genetics, 49(11):1567–1575, November 2017. PMID: 28991255.

[11] T. Li, J. Liu, J. Feng, Z. Liu, S. Liu, M. Zhang, Y. Zhang, Y. Hou, D. Wu, C. Li, Y. Chen, H. Chen, and X. Lu. Variation in the life history strategy underlies functional diversity of tumors. National Science Review, 8(nwaa124), February 2021.

[12] P. Hughes, D. Marshall, Y. Reid, H. Parkes, and C. Gelber. The costs of using unauthenticated, over-passaged cell lines: how much more data do we need? BioTechniques, 43(5):575, 577–578, 581–582 passim, November 2007. PMID: 18072586.

[13] N. Andor, B. T. Lau, C. Catalanotti, A. Sathe, M. A. Kubit, J. Chen, S. M. Grimes, C. J. Suarez, and H. P. Ji. Joint single cell DNA-seq and RNA-seq of gastric cancer cell lines reveals rules of in vitro evolution. NAR Genomics and Bioinformatics (in press), March 2020.

[14] Y. Liu, Y. Mi, T. Mueller, S. Kreibich, E. G. Williams, A. Van Drogen, C. Borel, M. Frank, P.-L. Germain, I. Bludau, M. Mehnert, M. Seifert, M. Emmenlauer, I. Sorg, F. Bezrukov, F. S. Bena, H. Zhou, C. Dehio, G. Testa, J. Saez-Rodriguez, S. E. Antonarakis, W.-D. Hardt, and R. Aebersold. Multi-omic measurements of heterogeneity in HeLa cells across laboratories. Nature Biotechnology, 37(3):314–322, March 2019.

[15] S. H. Selznick, M. L. Thatcher, K. S. Brown, and C. A. Haussler. Development and application of computer software for cell culture laboratory management. In Vitro Cellular & Developmental Biology. Animal, 37(1):55–61, January 2001. PMID: 11249207.

[16] S. G. Canfield, G. Jin, S. P. Palecek, and T. Sampsell. Software to improve transfer and reproducibility of cell culture methods. BioTechniques, 65(5):289–292, November 2018.

[17] S. Müller, A. Cho, S. J. Liu, D. A. Lim, and A. Diaz. CONICS integrates scRNA-seq with DNA sequencing to map gene expression to tumor sub-clones. Bioinformatics (Oxford, England), 34(18):3217–3219, September 2018. PMID: 29897414.

[18] A. Shukla, T. H. M. Nguyen, S. B. Moka, J. J. Ellis, J. P. Grady, H. Oey, A. S. Cristino, K. K. Khanna, D. P. Kroese, L. Krause, E. Dray, J. L. Fink, and P. H. G. Duijf. Chromosome arm aneuploidies shape tumour evolution and drug response. Nature Communications, 11(1):449, January 2020.

[19] S. Müller, E. D. Lullo, A. Bhaduri, B. Alvarado, G. Yagnik, G. Kohanbash, M. Aghi, and A. Diaz. A single-cell atlas of human glioblastoma reveals a single axis of phenotype in tumor-propagating cells. bioRxiv, page 377606, July 2018.

[20] L. Wang, J. Jung, H. Babikir, K. Shamardani, S. Jain, X. Feng, N. Gupta, S. Rosi, S. Chang, D. Raleigh, D. Solomon, J. J. Phillips, and A. A. Diaz. A single-cell atlas of glioblastoma evolution under therapy reveals cell-intrinsic and cell-extrinsic therapeutic targets. Nature Cancer, 3(12):1534–1552, December 2022.

[21] R. J. Beck and N. Andor. Local Adaptive Mapping of Karyotype Fitness Landscapes, July 2023. Pages: 2023.07.14.549079 Section: New Results.

[22] M. Baker. 1,500 scientists lift the lid on reproducibility. Nature, 533(7604):452–454, May 2016. PMID: 27225100.

[23] D. C. Minussi, M. D. Nicholson, H. Ye, A. Davis, K. Wang, T. Baker, M. Tarabichi, E. Sei, H. Du, M. Rabbani, C. Peng, M. Hu, S. Bai, Y.-W. Lin, A. Schalck, A. Multani, J. Ma, T. O. McDonald, A. Casasent, A. Barrera, H. Chen, B. Lim, B. Arun, F. Meric-Bernstam, P. Van Loo, F. Michor, and N. E. Navin. Breast tumours maintain a reservoir of subclonal diversity during expansion. Nature, 592(7853):302–308, April 2021. PMID: 33762732.